

# Replications: A Proposal to Increase their Visibility and Promote them

By LUCAS COFFMAN, MURIEL NIEDERLE AND ALISTAIR J. WILSON\*

A number of recent results in peer-reviewed academic work have ignited a debate over the fraction of results that are false positives. Published work in top Economics journals has discontinuities around  $p$ -values of 0.05, consistent with researcher data-mining (Brodeur et al., 2016). Moreover, in an attempt to replicate a set of recent experimental articles, Camerer et al. (2016) were able to successfully replicate two-thirds of the studies, which is less than the original  $p$ -values would suggest.

One response to such findings is to minimize flexibility over the econometric specifications available to researchers via registration and pre-analysis plans.<sup>1</sup> However, the sheer volume of research, coupled with the strong tendency to only finish and publish positive findings (Franco, Malhotra and Simonovits, 2014) leads to false positives even in the absence of somewhat questionable research practices (Ioannidis, 2005; Coffman and Niederle, 2015; Simonsohn, Nelson and Simmons, 2014). A solution is therefore required to try and separate what is true from what is not after a paper is disseminated: Replication.

While we discuss the dearth of replications, it is worth noting that replications do exist, though they often take forms that make them hard to find. Consider, for example, Niederle and Vesterlund (2007). Of the first two dozen independent replications (that is, without a coauthor of the initial study) none can easily be found when searching for “replication of Niederle and Vesterlund (2007).” This is because every one of these replications is part of (and buried within) a paper that has a much

larger scope than replicating Niederle and Vesterlund (2007) which is often only a side-result (see Niederle and Vesterlund, 2011; Niederle, 2016).<sup>2</sup> In addition, there have been replications that were simply not published since they were positive.<sup>3</sup>

In light of this fact that on the one hand replications are rare, and on the other hand the few that exist are hard to find (at least until the first survey papers) we make a two-pronged proposal for strengthening the incentivizes of new and better organizing existing replications in Economics. First we propose that top journals include one-page “replication reports.” One style of such reports could be new work that shows whether a specific published paper was replicated (or not)—such as the replication by Reuben, Sapienza and Zingales (2015) that was written up only much later after additional results were added. A second style of such reports consists of authors re-publishing an existing replication attempt that was already published elsewhere as part of a larger paper whose aim was not the replication of the specific published paper—such as the early replications of Niederle and Vesterlund (2007). Second, we propose a norm of citing replication work alongside the original. Whereas the current norm for citing might simply be “see Kessler and Roth (2012)”, our proposal would simply be to add, “replicated by Camerer et al. (2016).”

The aim of the proposal is to increase the visibility of replications but also increase the incentives for replications, since the

\* Coffman: Harvard University. Niederle: Stanford University, NBER and SIEPR. Wilson: University of Pittsburgh.

<sup>1</sup>See Olken (2015) and Coffman and Niederle (2015) for discussions of pre-analysis plans.

<sup>2</sup>The paper that can be easily found is the somewhat unique paper that does not replicate the results, which is also the only one that does have replication in the title.

<sup>3</sup>One of these found its way to be a paper by adding interesting results on external relevance (Reuben, Sapienza and Zingales, 2015).

costs can be substantial (see Hamermesh, 2017, for additional discussions). However, Coffman and Niederle (2015) suggest that even a small number of replications (less than a handful) can be very valuable to increase the posterior that a specific publication presents a robust result.

### The State of Replications

There are many ways in which we can replicate a result or provide some idea of its robustness. While what constitutes a replication or a robustness check may differ by field, we will broadly summarize these efforts when we use the term “replication” herein. Economics journals are reluctant to publish explicit replication attempts. Focusing only on studies where the main purpose was validating a previous empirical result, Duvendack, Jones and Reed (2017) find that in the past fifty years the *American Economic Review* has published 28 studies where the main purpose was validating an empirical result from a previously published study. Instead, the profession places much greater emphasis on novelty.

This is not to say that works focused exclusively on replicating others’ findings do not get published in prominent outlets. (Camerer et al., 2016) published their replication of eighteen prominent experimental papers in *Science*. The *Journal of Money Banking and Credit’s* macroeconomic replication exercise was documented in the *American Economic Review* (Dewald, Thursby and Anderson, 1986). Moreover, the publication of papers from both this session and a sister session focused on macroeconomic replications illustrate a desire to address the topic (Duvendack, Jones and Reed, 2017; Hoeffler, 2017; Anderson and Kichkha, 2017; Chang and Li, 2017).

However, while omnibus replications make clear contributions, more-comprehensive incentives for publishing and disseminating replication results would make clear improvements. Examining the AER’s centenary volume, Berry et al. (2017) find that after five years of publication only three out of every ten empirical

papers have a published replication.<sup>4,5</sup> Furthermore, only a minority of the remaining 70 percent had published robustness or extension work building on the original.<sup>6</sup>

In addition to the fact that replications are rare, is that in general these replications are hard to find. Pure replications, where the sole intent is to replicate and confirm others’ work seem to be quite rare outside of the larger-scale replication exercise cited above. Returning to the results of the analysis of replications by Berry et al. (2017), for the 29 percent of papers that did have published replications, all were embedded within more ambitious work, and took substantial search effort to weed out.

### A Proposal to Increase the Visibility of Replications

We would like to increase the visibility of extant replications and promote new replications through greater incentives to researchers. The ideal would be for a prestigious journal to introduce a short, new section. Given the venue for the current paper, the AER is a clear example, but the aim here would be for all top journals to follow suit.<sup>7</sup> In addition to *Shorter Papers*, *Comments* and *Replies*, there would be a *Replications* section. Publications within this section would be short: identifying the replicated paper, briefly describing any qualitative differences in procedures

<sup>4</sup>Hamermesh (2017) shows that for ten famous papers in labor, eight had been replicated at least once. Sunkhtankar (2017) finds about 8 percent of recent top development papers have been replicated.

<sup>5</sup>The profession does seem to do a better job at replicating important papers. Berry et al. find that the replication rate for papers with more than 50 published citations was nearly double the base rate at 60 percent. This is corroborated separately for development and labor by the replication rates found by Sunkhtankar and by Hamermesh.

<sup>6</sup>Berry et al. (2017) find that of the 1,546 published economics papers citing articles in the 2010 AER, only 52 were replications of said articles, where only seven of these replications were published in top-five journals.

<sup>7</sup>One prestigious first-mover can lead to changing norms. Consider the AER’s adoption of open-data policies in 2003, after which they were adopted by *Econometrica*, the *Review of Economic Studies* and the *Journal of Political Economy* (see McCullough, 2009, for a timeline).

(if any), and the main conclusion.<sup>8</sup> While the published replication paper would be an extended abstract, a more-extensive online appendix with data, code and formal tests would be linked for interested readers.

There are two reasons why this should increase the visibility and the availability of replications: First, we would allow authors who wrote and published a paper that builds on and extends a result from an important paper to re-publish aspects of their data that specifically replicates said important paper. While there are some costs in writing the replication paper, the incentive of an additional publication might be sufficient to encourage such an endeavor. Such an exercise would make trivial the task of finding “hidden” replications. Where the extensive coding in Berry et al. (2017), Hamermesh (2017), and Sunkhtankar (2017) took many hours of work to discover and code replications, our proposal would allow for discovery through a simple indexed search. Moreover, this process would de-bias the search for replications. Positive and negative replications would receive equal weight, where casual empiricism suggests that presently negative replications are much more visible as they have a much higher chance to be published.

There is a second reason why we think having a replication section at a prestigious journal like the AER would raise the number of replications. At times researchers try to replicate the results of new papers that are well published. For example, in many graduate programs, students still learning technical skills are often asked to replicate famous paper’s results. While these replications are occasionally published, this is more often the case when they fail to replicate the original paper, and even then publication often presents a big challenge, especially to young researchers.<sup>9</sup> For science

it would be equally valuable to know which papers are replicated, to get a better sense on how many papers fail to be replicated.

Once a desired number of replications were written, one can imagine that with many positive replications a coda summary article or a meta-analysis could be written—including all the authors of replications—to summarize the findings. Likewise, in the case of many failed replications. In the case of mixed results, more replications would be encouraged. That is with replications being visible, it would be more clear which papers still need replications, while others may be deemed to be robust given the large evidence of positive replications (for example, after a meta-analysis was written). Replication efforts could therefore be directed towards papers that are not yet replicated, as they would be published in the *AER*.

The second part of the proposal provides an incentive to produce replication work in the currency of our industry: citations. While it may be too costly for top journals to publish all replications, journals could encourage citation systems where citations to the original paper include citations to its replications. This would only cost a few extra lines, and would ensure that well-executed replications receive citations. A second benefit would be to the journal: If replications of especially highly cited papers are welcomed and published, and those replications are cited alongside the original article, this would ensure that the replication reports would have above average citations and hence not reduce the impact factor of the journal.

This proposal neglects many details. What constitutes a good and fair replication? How do we avoid unfair targeting of the projects that get replicated? Which replications merit a report in top journals?

<sup>8</sup>One potential issue here is the independence of replications: whether authors of the original study can “replicate” themselves. While we might, as a profession, not want to “lose” informative replications, clearly the onus should be on independent attempts.

<sup>9</sup>For example, the replication of Niederle and Vesterlund (2007) that is now part of Reuben, Sapienza and Zingales (2015) while being one of the early replications

(it was run in the fall of 2006) would probably not have seen the light of day, as it was a positive replication. Via private correspondence with us, one of the authors of that project told us that they had discussed writing a short note on the replication but concluded “it would not be worth it because the payoff is so low.” It was only recovered by adding substantial new data which makes the replication a small part of a larger paper.

Who should be making editorial and reviewing decisions? Given the space constraints of this article, we simply put forward a non-controversial first step, one that we can agree on. But we recognize the difficulty in answering these questions (among many others) more fully.

Finally, we suggest these policies need to come from the top down. Replication work, even if published, needs visibility to affect awareness; it needs the visibility afforded by top journals. Further, the norms of citing replication work will only be solidified if it comes in well-read papers. This can only happen if done by well-published authors and enforced by editors at top journals.<sup>10</sup>

Many researchers are concerned by the recent upheaval in psychology determining that most papers cannot be replicated. Clearly the issue of replicability is a source for substantial concern in economics as well. The *Replication Network* (<http://replicationnetwork.com>) whose aim is to promote replications and to ultimately have economics journals publish those replications, had been endorsed by 340 economists at the time of writing. While we believe the economics profession is not doing too badly on the dimension of replicability, this is an area that economists can be leaders in designing better mechanisms for promulgating academic research.

## REFERENCES

- Anderson, Richard, and Areerat Kichkha.** 2017. “Replication versus Meta-Analysis in Economics: Where Do We Stand 30 Years after Dewald, Thursby and Anderson?” *American Economic Review*, 107.
- Berry, James, Lucas C. Coffman, Rania Gihleb, Douglas Hanley, and Alistair J. Wilson.** 2017. “Assessing the Rate of Replication in Economics.” *American Economic Review*, 107.
- <sup>10</sup>As discussed in Nosek, Spies and Motyl (2012), new replication or negative-result journals are a non-starter. Without the prestige of a large professional association or a high-impact journal to begin with, such journals are doomed to failure.
- Brodeur, Abel, Mathias Lé, Marc Sangnier, and Yanos Zylberberg.** 2016. “Star wars: The empirics strike back.” *American Economic Journal: Applied Economics*, 8(1): 1–32.
- Camerer, Colin F., Anna Dreber, Eskil Forsell, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Johan Almenberg, Adam Altmejd, Taizan Chan, et al.** 2016. “Evaluating replicability of laboratory experiments in economics.” *Science*, 351(6280): 1433–1436.
- Chang, Andrew C., and Andrew Li.** 2017. “Is Economics Research Replicable? Sixty Published Papers from Thirteen Journals Say ‘Usually Not’.” *American Economic Review*, 107.
- Coffman, Lucas C., and Muriel Niederle.** 2015. “Pre-analysis plans have limited upside, especially where replications are feasible.” *Journal of Economic Perspectives*, 29(3): 81–97.
- Dewald, William G., Jerry G. Thursby, and Richard G. Anderson.** 1986. “Replication in empirical economics: The Journal of Money, Credit and Banking project.” *American Economic Review*, 587–603.
- Duvendack, Maren, Richard Jones, and Robert Reed.** 2017. “What is Meant by ‘Replication’ and Why Does It Encounter Such Resistance in Economics?” *American Economic Review*, 107.
- Franco, Annie, Neil Malhotra, and Gabor Simonovits.** 2014. “Publication bias in the social sciences: Unlocking the file drawer.” *Science*, 345(6203): 1502–1505.
- Hamermesh, Daniel S.** 2017. “Replication in Labor Economics: Evidence from Data and What It Suggests.” *American Economic Review*, 107.
- Höeffler, Jan H.** 2017. “Replication and Economics Journal Policies.” *American Economic Review*, 107.

- Ioannidis, John P.A.** 2005. "Why most published research findings are false." *PLoS Med*, 2(8): e124.
- Kessler, Judd B., and Alvin E. Roth.** 2012. "Organ allocation policy and the decision to donate." *American Economic Review*, 102(5): 2018–2047.
- McCullough, Bruce D.** 2009. "Open access economics journals and the market for reproducible economic research." *Economic Analysis and Policy*, 39(1): 117–126.
- Niederle, Muriel.** 2016. "Gender." In *Handbook of Experimental Economics*. Vol. 2, , ed. John Kagel and Alvin E. Roth. Princeton University Press.
- Niederle, Muriel, and Lise Vesterlund.** 2007. "Do women shy away from competition? Do men compete too much?" *Quarterly Journal of Economics*, 1067–1101.
- Niederle, Muriel, and Lise Vesterlund.** 2011. "Gender and competition." *Annual Review of Economics*, 3(1): 601–630.
- Nosek, Brian A., Jeffrey R. Spies, and Matt Motyl.** 2012. "Scientific utopia II. Restructuring incentives and practices to promote truth over publishability." *Perspectives on Psychological Science*, 7(6): 615–631.
- Olken, Benjamin A.** 2015. "Promises and perils of pre-analysis plans." *Journal of Economic Perspectives*, 29(3): 61–80.
- Reuben, Ernesto, Paola Sapienza, and Luigi Zingales.** 2015. "Taste for competition and the gender gap among young business professionals." National Bureau of Economic Research Working Paper 21695.
- Simonsohn, Uri, Leif D. Nelson, and Joseph P. Simmons.** 2014. "P-curve: a key to the file-drawer." *Journal of Experimental Psychology: General*, 143(2): 534.
- Sunkhtankar, Sandip.** 2017. "Replications in Development." *American Economic Review*, 107.